

一文读懂 Sora2 核心点

核心观点

Sora 2 以“为产品定义功能”的产品思维，构建了从拉新、留存到促活的产品闭环，3 天登顶 iOS 应用榜，开启了 P、B、C 三端共振的千亿级 AI 视频生成赛道。据测算，AI 视频生成市场中期空间将达到 763 亿元，长期将达到 1554 亿元。同时，根据我们正文的测算依据，保守估计 Sora APP 每日仅推理成本便高达 1400 万美元，年化成本超过 51.2 亿美元，预计将带来持续的算力需求。建议关注两大主线，AI 应用推荐以阿里为代表的互联网大厂生态伙伴、Pre-AI 环节和垂直场景；算力环节推荐算力和端侧 AI 相关标的。

要点

“为产品定义功能”，Sora 2 围绕社交打造全链路能力

Sora 2 三天内登顶美国 iOS 应用榜，证明了将复杂技术封装于极致简洁的创作流程，并嫁接“Cameo”（数字分身）与“Remix”（二创）等强社交功能，能够有效引爆大众市场的病毒式传播。Sora 2 将 AI 工具转变为社交和娱乐产品，成功构建了从拉新、留存到促活的产品闭环，为 AI 应用打造了产品化模板。

P 端+B 端+C 端共振，AI 视频生成千亿市场可期

P 端用户付费意愿强，我们预计全球 P 端市场中期空间为 262 亿元，长期将达 888 亿元；B 端市场场景应用广，核心在于对传统制作流程的成本替代，我们预计中期市场可达 501 亿元，长期增长至 666 亿元；预计 Sora APP C 端可能将效仿短视频应用模式，通过 AI 原生内容社区汇聚流量变现。综合来看，AI 视频生成赛道中期市场总空间预计为 763 亿元，长期将达到 1554 亿元。

Sora APP 将带来年化 50 亿美元以上推理算力需求

根据 Transformer 架构和 DiT 架构论文，我们保守估计 Sora APP 每日仅推理成本便高达 1400 万美元，年化成本超过 51.2 亿美元。巨大推理算力需求与 OpenAI 近期和 Nvidia/AMD 等芯片巨头达成的战略合作展现了同样的趋势，凸显了算力在未来 AI 竞争格局中的决定性地位。

风险提示：AI 产业商业化落地不及预期；市场竞争风险；政策风险；地缘政治风险。

计算机

维持

强于大市

应瑛

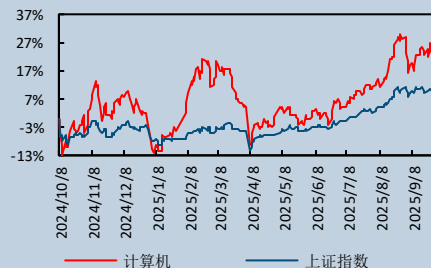
yingying@csc.com.cn

SAC 编号:S1440521100010

SFC 编号:BWB917

发布日期：2025 年 10 月 08 日

市场表现



相关研究报告

- | | |
|----------|--|
| 25.09.21 | 【中信建投计算机】:周报 25 年第 35 期:云栖大会将近,看多阿里 AI 和端侧 AI |
| 25.09.17 | 【中信建投人工智能】:Oracle 云订单暴涨,算力景气度持续上行 |
| 25.09.14 | 【中信建投计算机】:周报 25 年第 34 期:Oracle 验证大厂资本开支向上,看好阿里 AI 全栈布局 |
| 25.09.07 | 【中信建投计算机】:周报 25 年第 33 期:海外大厂资本开支目标明确,国内模型加速迭代 |
| 25.09.05 | 【中信建投人工智能】:液冷散热系列报告一:热界面材料——搭建芯片等电子元器件的高速散热通道 |

目录

一、Sora 2 为什么能够爆火？	1
1.1 视频生成技术发展历程：从范式探索到架构收敛.....	1
1.2 Sora 2 爆火：产品工程化+社媒裂变.....	3
二、如何才能使用 Sora 2？	4
2.1 邀请制构建社交生态，核心创作流程极致简化.....	4
2.2 Sora 2 围绕社交链路打造重点功能.....	6
三、Sora 2 视频生成的效果如何？	8
3.1 文生视频（武侠）	9
3.2 文生视频（悬疑）	10
3.3 图生视频（喜剧、卡通）	12
3.4 图生视频（科幻）	14
四、AI 视频生成市场空间有几何？	16
五、Sora 2 的 Token 成本有多少？	18
六、投资建议	20
风险分析	21

一、Sora 2 为什么能够爆火？

1.1 视频生成技术发展历程：从范式探索到架构收敛

视频生成技术是一条追求更高生成质量、更长内容时长与更强语义可控性的演进之路。视频生成承接图像生成技术，早期以生成对抗网络（GAN）和变分自编码器（VAE）为代表，通过生成连续图像帧来构建视频，但受限于稳定性与扩展性瓶颈，往往只能生成分辨率低、时长短的片段。随着技术不断演进，业内分化出两条并行路线，Transformer 架构凭借其高效处理视频帧之间时序关系的能力展现出巨大扩展潜力；扩散模型（Diffusion Model）则通过从随机噪声中逐步还原画面的“去噪”范式，在生成质量上取得里程碑式突破。2022 年 12 月，融合二者优点的 Diffusion Transformer（DiT）架构论文发表，经由 OpenAI Sora 产品侧验证后，逐步确立了其作为视频生成领域主流技术范式的地位。

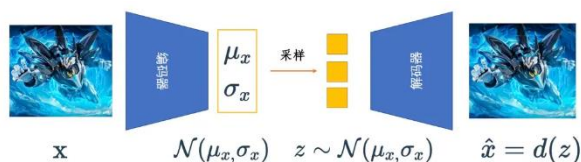
图 1：视频生成技术主要发展历程

	探索期	早期	增长期	未来方向
	GAN	Transformer	Diffusion Model	Diffusion + Transformer
技术节点	2014年GAN发表，2016年左右开始用于视频生成 2013年VAE出现，可以通过编码器和解码器的机制学习数据的潜在表示	2017年Transformer发表，逐步向各个领域渗透，早期在视频生成领域也有尝试 2021年Google发布Video Vision Transformer	2020年 DDPM 论文发表 2021年 Latent Diffusion Models 论文发表 2022年 Video Diffusion Model 论文发表	2022年12月Diffusion Transformer论文发表 2024年2月OpenAI发布Sora，取得惊艳的生成效果，验证了Diffusion和Transformer结合的有效性
案例	- Text2Filter、TGANs-C、VGAN、TGAN、VideoGPT、MoCoGAN、DVD-GAN、DIGAN - 应用范围有限（只能总特定生成，例如数字），生成分辨率低、生成时长极短	- GODIVA、VideoGPT、Phenaki、Cog Video、NUWA - 早期基于Transformer的视频生成效果有限，在当时的成本比较高	- Stable Video Diffusion、Make-A-Video、Runway Gen1/2、Video LDM - 由于SD的开源激发社区创新，推动扩散模型在视觉生成领域占据主导地位	- Sora、Videopoet、W.A.L.T - 由于领军企业的示范效应，Diffusion Transformer已经成为视频生成领域的核心路线，领域内其他玩家都开始聚焦DiT并取得进展 - 目前业界对于Sora的架构大多是猜想，从架构大方向上接近，但具体细节上不同模型、产品各异
特点	- 生成多样性差：生成器-判别器对抗的模型架构决定生成效果逼真但多样性、创造性差，很难泛化 - 生成速度快：GAN可以一次出图，无需做多步推理	- 可扩展性强：Scale up路线上目前最好的架构 - 生成速度慢、开销大：需要把图片进行分割再转换成超长序列，计算量随分辨率增加平方级增加，复杂度高	- 生成质量高：生成效果好，细节丰富清晰，训练要比GAN稳定 - 显存开销大：视频生成有大量相关依赖关系的图片同时生成，对计算显存有极高要求	- 可扩展性强：基于扩散模型的视频模型，核心的构件是卷积神经网络U-Net，随着模型规模scale up增益会放缓甚至消失。DiT（Diffusion Transformer）将传统扩散模型的U-Net替换为Transformer，从而能够实现更强的可拓展性 - 长期来看语言模型路线¹可成为主要路线，可以将多种模型融合到一起（由于架构的成熟和优良的扩展性）

数据来源：量子位智库，中信建投

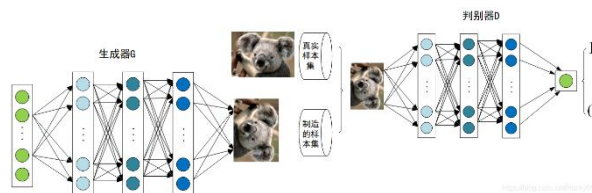
第一阶段（2017 年以前）：图像拼接。VAE 于 2013 年提出，采用编码器-解码器架构，先将图像压缩至低维度的潜在空间，再由解码器从空间中采样并重建图像，从而学习数据的核心分布。VAE 的目标函数倾向于生成模糊平滑的图像，难以捕捉高频细节。2014 年发表的 GAN 则引入了对抗性训练机制，由生成器从随机噪声合成图像，再由判别器进行真伪判别，二者竞争下生成器能够产出更真实的图像，判别器也能够更好区分图像真实性。然而，由于 GAN 的目标函数没有显式地鼓励生成多样性，生成器在优化过程中往往会找到局部最优解，导致模式崩溃问题。此外，基于 VAE 和 GAN 的视频生成核心在于将二者生成的图片帧进行拼接，缺乏对长距离时序关系的有效建模，使得视频只能在短时间连贯，无法维持逻辑与内容的一致性。

图 2：VAE 模型示意图



资料来源：《Auto-Encoding Variational Bayes》，腾讯技术工程，中信建投

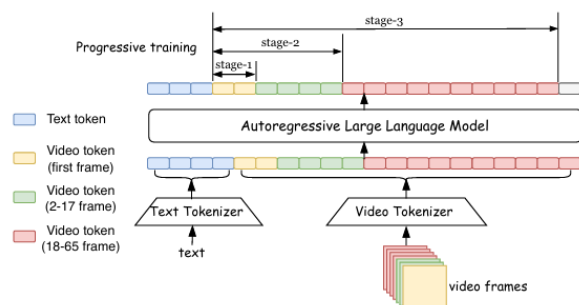
图 3：GAN 模型示意图



资料来源：《Generative Adversarial Nets》，CSDN，中信建投

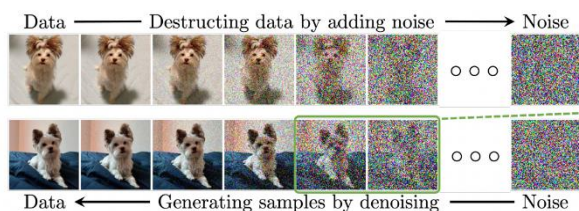
第二阶段（2017-2023 年）：多路径探索。为解决前述缺乏时空一致性及图像分辨率不足等问题，业内多种技术路径并行探索。其中，自回归路径将视频帧视为离散的 Token 序列，并逐个 Token 地预测生成后续内容。为有效建模视频所需的长距离时序依赖，自回归路径通常采用 Transformer 等序列模型，其一定程度维系了视频生成的时序连贯性，但串行的生成机制也带来了推理速度慢、误差易于累积等挑战。扩散模型路径则采用并行“去噪”方式生成整个视频片段，极大地提升了生成画面的质量与稳定性。扩散模型路径早期普遍采用 U-Net 网络架构，催生了以 Runway Gen-2 为代表的系列产品，并迅速成为市场主流，但其在模型规模进一步扩大时逐渐暴露出扩展性上的瓶颈。

图 4：一种基于自回归的视频生成模型示意图



资料来源：《Loong: Generating Minute-level Long Videos with Autoregressive Language Models》，中信建投

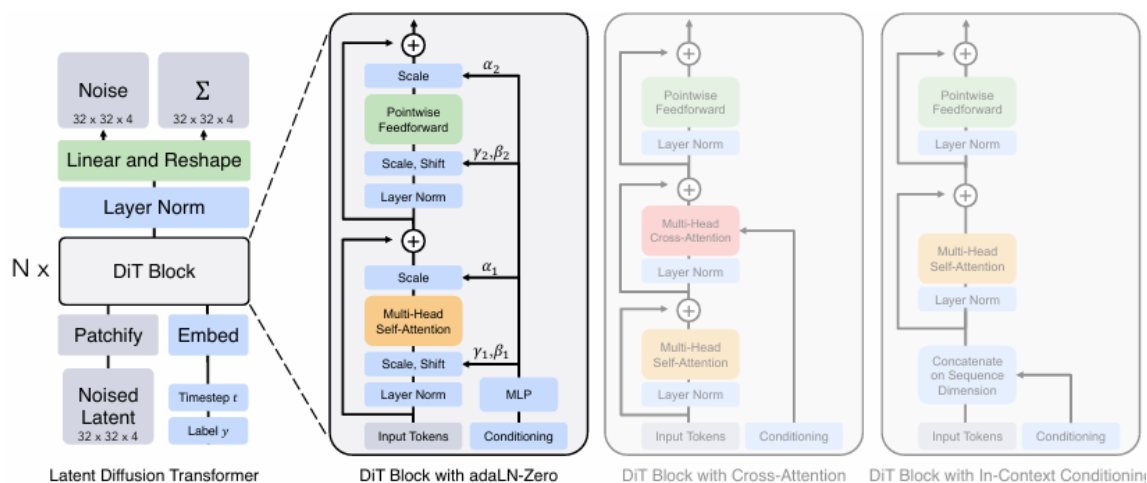
图 5：扩散模型向前加噪和向后去噪过程



资料来源：阿里云开发者社区，中信建投

第三阶段（2023 年至今）：DiT 引领主流技术趋势。2022 年 12 月，《Scalable Diffusion Models with Transformers》发表，其核心思想在于用 Transformer 架构取代主流扩散模型路径中的 U-Net 骨干网络，并证明了这种新架构（DiT）拥有卓越的扩展性。2024 年 2 月，OpenAI 发布的 Sora 进一步在产品侧验证了 DiT 的可行性，Sora 的惊艳效果亦迅速推动 DiT 成为行业公认的主流范式。此后，市场上涌现出系列采用或对标 DiT 架构的模型，如包括生数科技的 Vidu、Google 推出的 Veo 等，标志视频生成领域已进入 DiT 架构主导的新阶段。

图 6：DiT 架构示意图

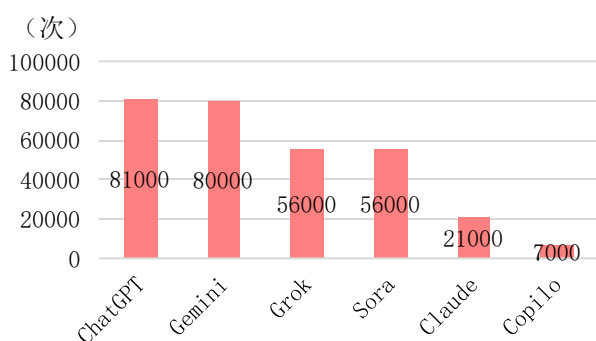


数据来源：《Scalable Diffusion Models with Transformers》，中信建投

1.2 Sora 2 爆火：产品工程化+社媒裂变

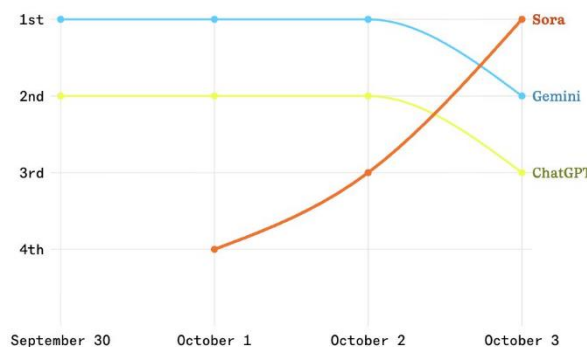
Sora 2 三天登顶美国 iOS 应用榜。据应用情报提供商 Appfigures 数据，Sora 的 iOS 应用程序首发下载量 5.6 万，并在上线的两天内总计获得了 16.4 万次下载。尽管 Sora 的首日下载量落后于 ChatGPT 和 Gemini，但考虑其在仅限美国和加拿大地区上线并采用邀请制的情况，其用户偏好可能更强。当地时间 10 月 3 日（周五），Sora 在上线第三天登顶苹果美国应用商店 App Store 的免费应用榜第一名，超越了 OpenAI 的 ChatGPT 和谷歌的 Gemini。

图 7：部分 AI 应用首日下载量



资料来源：Appfigures, Medium, 中信建投

图 8：10 月 3 日 Sora 登顶美国地区应用商店免费榜



资料来源：Appfigures, 新智元, 中信建投

尽管 Sora 2 尚未发布技术报告，但我们根据现有信息推测其并未在技术上明显突破，而是训练数据、产品工程化等方面做出优化：

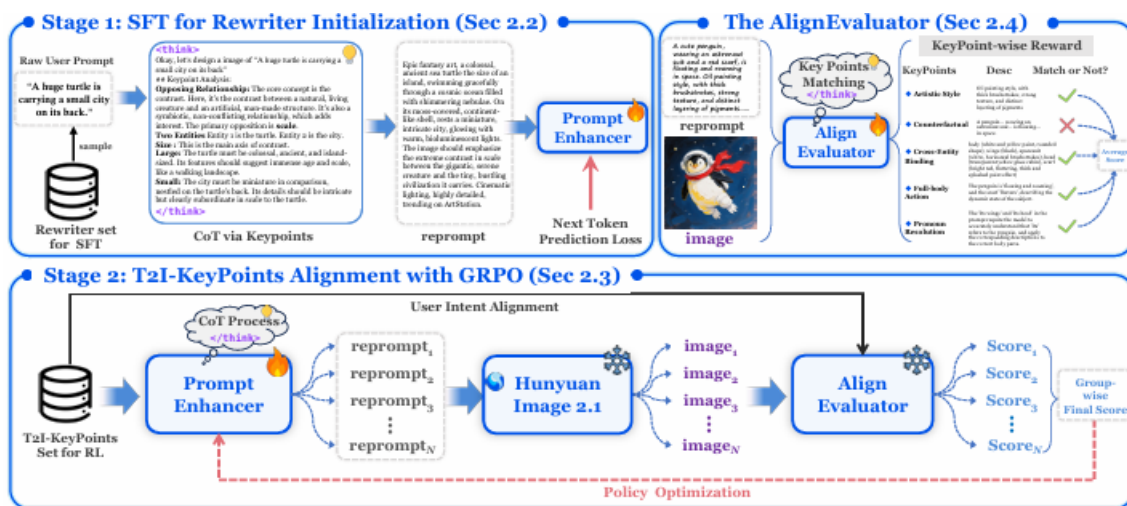
大规模数据训练：据 OpenAI 官网 Sora 2 展示页面表示，Sora 2 的一个重要里程碑是掌握大规模视频数据的预训练和后训练方法。OpenAI 称，尽管 Sora 2 模型仍存在许多错误，但其证实了进一步扩大视频数据上的神经网络将使模型更接近模拟现实。

指令遵循与音画同步：Sora 2 在可控性方面也实现了较大突破，其能够执行跨越多个镜头的复杂指令，同

时精准地保留世界状态。此外，作为通用的视频音频生成系统，Sora 2 能够创建具有高度真实感的复杂背景音景、语音和音效。Sand.ai 创始人曹越在接受极客公园访谈时表示：“从纯技术角度看，把声音做好，可能不像把视频或语言做好那么难。但从产品和用户视角看，声音是决定性的。正是因为 Sora 能让音画同步输出，才使得普通用户生成的视频具备了直接的可消费性，越过了 C 端产品普及的临界点。”

提示词增强/重写：尽管 Sora2 是视频模型，但其亦可以解决 LLM 基准测试中的问题。Epoch AI 在小部分 GPQA 问题上测试了 Sora 2，其得分为 55%（GPT-5 得分为 72%）。据 Epoch AI 推测，Sora 2 用户的提示可能在视频生成之前被 LLM 重写，即 LLM 层可能会先解决问题，然后将解决方案明确地包含在重写的提示中。实际上，混元亦于 9 月开源了 PromptEnhancer 技术架构，仅通过“思维链（CoT）提示重写”的思路使 AI 生图的对齐精度大幅提升，在抽象关系理解、数值约束等复杂场景中，准确率甚至能提升 17% 以上。

图 9: PromptEnhancer 技术架构



数据来源:《An overview of the training framework for PromptEnhancer》, 中信建投

产品侧重视 AI+社交则是 Sora 2 爆火的另一重要因素。相较于此前视频生成模型追求高分辨率的输出，Sora 2 仅生成 360p 画面并免费供受邀用户使用，大幅降低了用户门槛。Cameo 功能则支持用户制作保留面部微表情与声纹特征的数字分身，并可将分身植入自己或朋友生成的视频场景。考虑到邀请码的裂变机制（受邀用户也会生成 4 个新的邀请码）本身强调用户更可能将产品分享给熟悉的朋友，更进一步为 Cameo 功能提供了基础。通过朋友共创+私域传播，Sora 2 成功打造了“AI+社交”的产品氛围，加速产品爆火节奏。

Sora 2 的爆火标志着 AI 应用竞争进入了新阶段——单纯的技术参数比拼之下，更重要的是将技术能力无缝封装进优秀的产品体验、并设计出能够自我驱动的疾病式传播循环，才是引爆大众市场的关键。

二、如何才能使用 Sora 2？

“好的产品做减法”。Sora 2 的使用流程看似简洁，但其背后每一步都充满了深思熟虑的产品设计，旨在最大化地降低创作门槛、激发社交欲望并构建用户粘性。

2.1 邀请制构建社交生态，核心创作流程极致简化

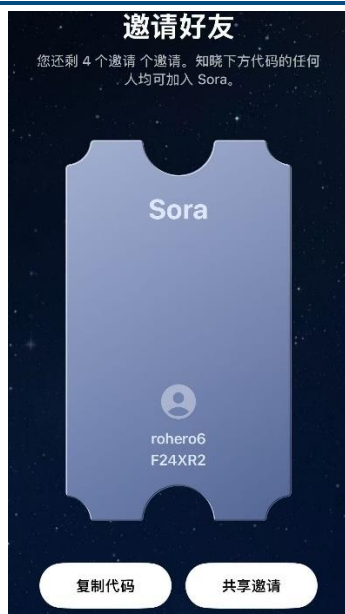
邀请制构建社交生态，移动端优先强调娱乐属性。Sora APP 发布之初仅开放了美国和加拿大 iOS 端下载

请务必阅读正文之后的免责条款和声明。

（截至 10 月 6 日，七麦数据显示排名的地区还包括日本、韩国和瑙鲁），网页端及安卓版则在后续逐步适配。完成 APP 下载后，用户可通过 OpenAI 账户直接登录，但需填写邀请码才能进入产品界面，获得邀请的用户则会得到 4 个新的邀请码。

Sora 2 的准入模式直接体现 OpenAI 对 C 端 AI+社交的重视：1) **移动端优先**：Sora 2 初期仅上线移动端，直接对标 TikTok 等短视频平台，相较过去视频生成工具用于降本增效，更强调“社交娱乐”而非“生产力工具”的初始定位。2) **生态绑定**：Sora APP 与 ChatGPT 账户直接关联，不仅为 OpenAI 产品生态构建了新吸引点，更将 ChatGPT 7 亿周活用户圈定为 Sora 的潜在客群。Sora APP 为 ChatGPT Pro 账户提供 Sora Pro 能力也展现了其为 ChatGPT 付费订阅导流的目标。3) **邀请制**：邀请制不仅能够控制初期高昂的算力成本，还能够营造“一码难求”的社交热度（格隆汇：首日邀请码最高被炒作至数十美元），激发用户渴望。此外，正如第一问提到的，邀请码的裂变首先会通过熟人链条传播，既保证初期用户群质量，又为社区交互提供了基础。

图 10：Sora APP 通过邀请码机制传播



资料来源：Sora APP，中信建投

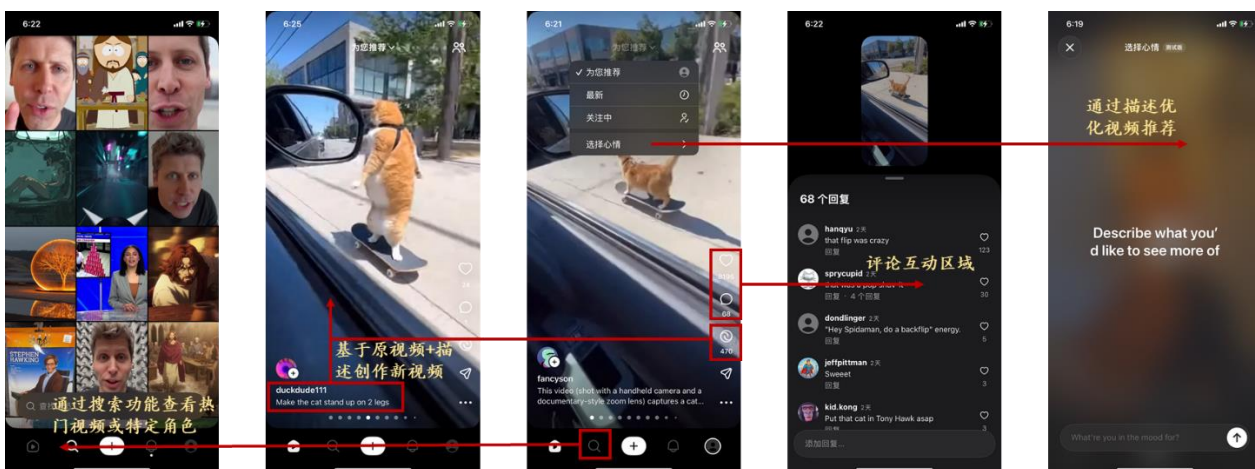
图 11：Sora APP 初期仅开放 Apple Store 下载



资料来源：AppleStore，中信建投

Sora 2 在 y 轴滑动切换作品的基础上为 Cameo 功能新增 x 轴。Sora 2 产品设计与短视频产品高度相似，主页为类似 TikTok 的单列推荐信息流，并提供点赞、评论、转发、搜索等基础功能，用户还可以通过自然语言描述优化推荐内容。除了常规的 Y 轴上下滑动切换不同作品外，Sora 2 创新性地引入了 X 轴左右滑动，这一设计专为 Cameo 和 Remix 功能服务，用户可以通过左右滑动，浏览不同创作者对同一原始作品的“二创”版本，极大地增强了内容的探索性和社区的互动性。

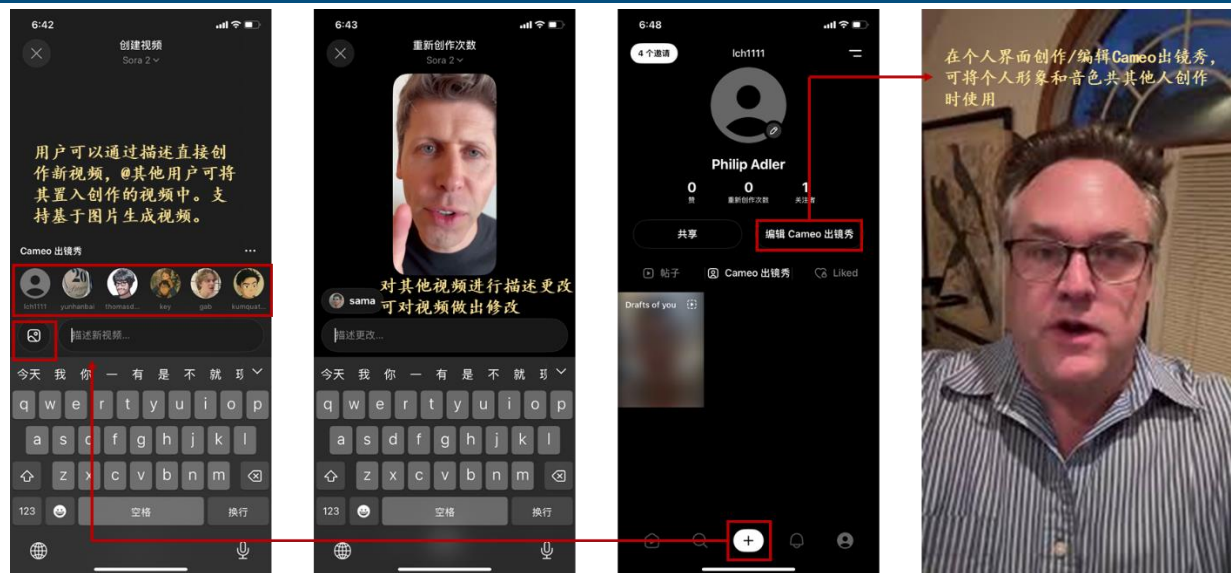
图 12：Sora APP 浏览主界面



数据来源：Sora APP，中信建投

Sora 2 的创作界面贯彻了“做减法”的原则。Sora2 摒弃了其他 AI 视频工具复杂的运镜、帧率等专业参数，取而代之的是一个极致简洁的文本输入框。用户只需用自然语言描述想法，即可“一句话生成视频”，功能上仅保留了图生视频和调用 Cameo 这两个核心选项。这种极致的简化，精准地契合了短视频受众追求简单、有趣而非复杂的创作心理，为产品的病毒式传播扫清了障碍。

图 13：Sora APP 创作主界面



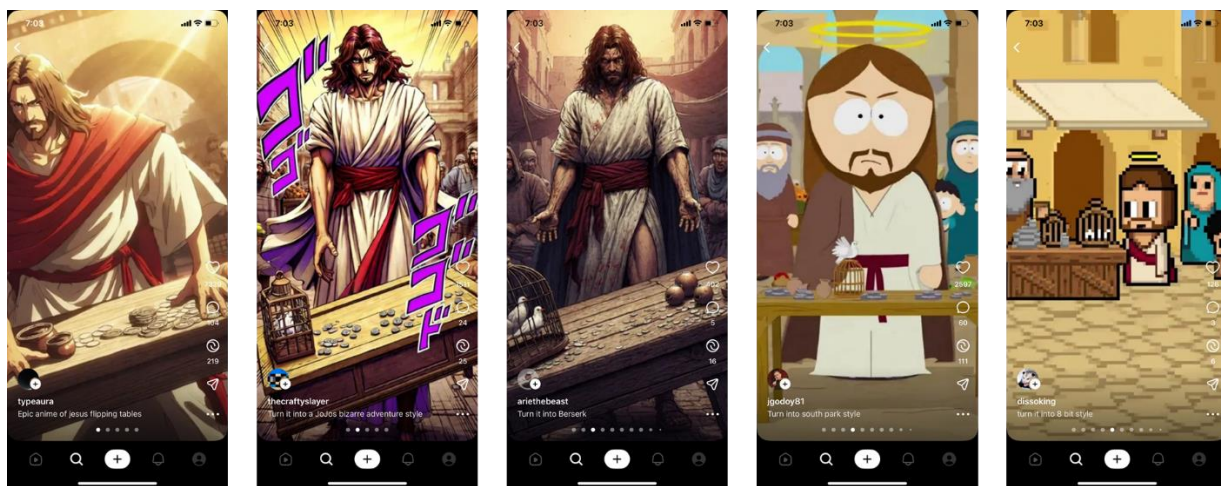
数据来源：Sora APP，中信建投

整体而言，Sora 2 的浏览和创作界面都相当简洁，其中浏览界面上下左右滑动的操作方式符合用户习惯，核心创作流程更是去掉了复杂的功能点。上述改动与短视频受众追求简单、有趣而非复杂、细致的目标相契合，为后续广泛传播提供了基础。

2.2 Sora 2 围绕社交链路打造重点功能

Remix: Sora 2 构建 UGC（用户生成内容）生态和社交网络的基石。 Remix 功能允许用户在浏览信息流时，选择任意一个感兴趣的视频作为“模板”进行二次创作。用户可以保留原视频的构图、风格或主体，仅修改部分提示词来改变剧情走向或替换关键元素，极大程度降低原创门槛。尤其当一个“Meme（迷因/梗）”诞生并病毒式传播时，Remix 将形成该 Meme 的内容流，大幅丰富平台的内容生态，从而将 Sora 2 从生成工具转变为内容社区，提升用户粘性。

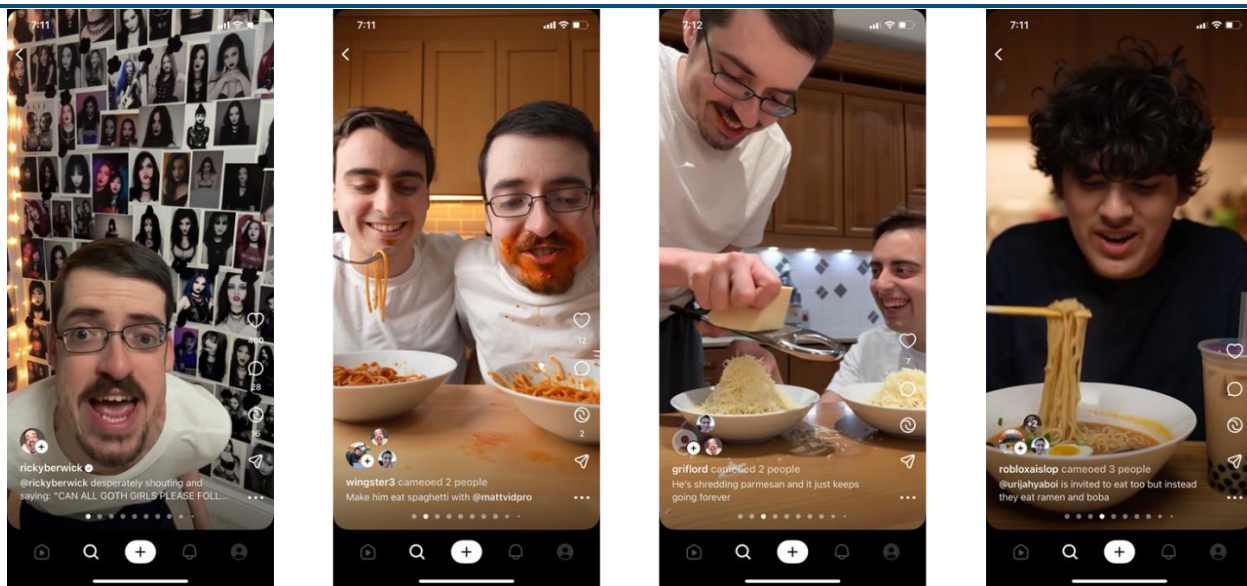
图 14: Sora APP Remix 功能演示



数据来源: Sora APP, 中信建投

Cameo（数字分身）：结合 Remix 打造社交裂变可能性。通过@不同好友的 Cameo 将其加入 AI 生成内容，虚拟的视频内容和真实的社交关系产生了交互，从而产生了情绪价值。具体而言，Cameo 将 AI 工具转化社交语言和娱乐方式，强化用户在熟人圈中分享和互动的欲望，形成了“拉新-留存-促活”的完美产品闭环。

图 15: Sora APP Cameo+Remix 功能演示



数据来源: Sora APP, 中信建投

为产品定义功能，而非为模型定义产品。“模型及产品（MaaS）”、“AI 吞噬软件”等论调都是基于强大

的模型能够解决各种软件功能的思路，而 Sora 2 给出了新的解法。除了 Remix 和 Cameo 两大核心功能，Sora 2 的其他优化同样反映了其服务社交链路的核心思路。1) **原生音画同步大幅改善了用户体验**。带声音的视频无需二次加工即可做为完整内容直接分享，极大地缩短了从创作到传播的路径。相较于 Veo3 的音画同步，Sora 2 的 Cameo 功能可直接将他人音色嵌入视频生成，更进一步强化了社交属性。2) **低画质和低可控性带来低门槛**。Sora 2 的视频牺牲了高画质和专业参数控制，大幅降低了生成视频所需的算力成本，为初期的免费模式提供了可能；极致简化的操作更是让任何用户都能在几秒钟内上手，确保产品有最广的 MAU 和裂变能力。

三、Sora 2 视频生成的效果如何？

随着 AI 视频生成技术不断发展，各类文生/图生视频产品百花齐放，且陆续开始商业化探索。截至目前，主流 AI 视频生成产品大多可生成 5-10s，1080p 视频。商业化多采取 C 端按会员订阅赠送积分以供生成视频，B 端按 API 调用付费的模式。

表 1:部分 AI 视频生成产品情况

模型/产品	开发商	核心优势与特点	最高输出规格	商业化
Sora 2	OpenAI	音画同步生成、物理真实感强、Cameos（数字分身）、有独立的社交 App。	免费：360p，10 秒；Pro 版：约 480p，15 秒；API 调用支持生成 1024p。	通过邀请码可以免费使用，ChatGPT Pro 会员（\$200/月）可以享受更强的 Pro 能力；Sora 2/Sora 2 Pro 生成 720p 分辨率的视频每秒分别 \$0.1/\$0.3；Sora 2 Pro 生成 1024p 分辨率的视频每秒 \$0.5。
Veo 3	Google (DeepMind)	运动一致性好、电影级质感、原生音频、与 Google 生态深度整合。	1080p，24 帧，8 秒。	Google AI Pro（\$20/月）每天可以生成 3 个视频；带音频的标准版模型 API 调用价格 \$0.4/秒。
Runway (Gen-4)	Runway ML	专业级编辑与控制工具（导演模式）、API 完善、迭代速度快、支持 4K 上采样。	默认 720p，可升至 4K，24 帧，8 秒。	基础/标准/高级会员每月分别 15/35/95 美元，分别赠送 625/2250/2250 积分（高级会员可使用更好的模型），生成 5 秒视频约消耗 60 积分。
Kling 2.1	快手	复杂和大幅度运动场景处理出色、对中国文化元素理解深入。	1080p，10 秒，24 帧。	通过灵感值生成视频，普通用户登录送的灵感值可用于免费生成视频。黄金/铂金/钻石/黑金会员（不考虑优惠）每月分别 58/234/586/1314 元，对应每月 660/3000/8000/26000 灵感值，生成单个标准视频约消耗 20 灵感值。
Seedance 1.0	字节跳动	单图驱动角色动画，尤其擅长生成连贯、富有节奏感的舞蹈视频。	1080p，10 秒，24 帧。	C 端接入即梦、豆包等产品免费提供；B 端通过火山引擎平台为提供服务，API 调用 0.015 元/千 Tokens，生成一条 5 秒的 1080p 视频需 3.67 元。
Vidu Q1	生数科技	长时程（16s+）生成能力、物理世界模拟、对中国元素理解深刻、DiT 架构。	1080p，5 秒，24 帧。	C 端消耗积分生成视频，普通用户可以通过登录赠送积分生成视频，标准版/专业版/旗舰版每月分别 59/219/629 元，每月分别赠送 800/4000/8000 积分，生成单个视频约消耗 4 积分；B 端 API 为主，同样采用积分模式，1 积分 0.3125 元，生成 5 秒的 1080p 视频需 8 积分（错峰 4 积分）。
Wan 2.2	阿里巴巴	单图+音频驱动的数字人，口型精准、表情生动自然，擅长生成唱歌/说话视频。	1080p，5 秒，30 帧。	C 端消耗灵感值生成视频，普通用户可以通过登录赠送灵感值生成视频，标准会员/高级会员每月分别 72/288 元（包年 5 折），每月分别赠送 300/1200 灵感值，单个视频约消耗 5 灵感值；B 端通过 API 提供服务，wan2.2 专业版生成 1080p 视频 0.7 元/秒。
PixVerse V5	爱诗科技	迭代速度快、社区驱动、创意风格化效果突出、易于使用。	1080p，8 秒，24 帧。	C 端消耗积分生成视频，普通用户可以通过登录赠送积分生成视频，标准版/Pro 版/Premium 版每月分别 \$10/30/60，每月分别赠送 1200/6000/15000 积分，720p/5 秒视频消耗 5 积分，4K/30 秒视频需 85 积分。针对企业用户提供定制化服务包，包含 API 接入、私有化部署和专属模型训练，标准企业套餐起价为 \$500/月，包含 30000 积分和专属技术支持。大型机构可选择年度合约方案，享受 15% 的批量折扣及数据安全合规保障。

资料来源：OpenAI，Runway，Google，可灵 AI，阿里万相，字节跳动，生数科技，爱诗科技，中信建投

为了比较 Sora 2 AI 视频生成的效果，我们采用久谦科技咨询在文生视频（武侠）、文生视频（悬疑）、图生视频（喜剧、卡通）、图生视频（科幻）场景的测评结果，并用同样的提示词要求 Sora 2 生成对应场景的视频进行对比。其中，每次模型只取第一次测试结果，不进行多次测试。

请务必阅读正文之后的免责条款和声明。

总结：尽管 Sora 2 牺牲了较多视频控制的相关参数，但在文生视频（武侠）、文生视频（悬疑）、图生视频（喜剧、卡通）三大场景中均实现了第一梯队甚至领先的水平，仅在图生视频（科幻）场景测试中相对较弱。实际上，四次生成中 Sora 2 运镜表现均相对一般，场景切换也略显突兀，因此在增加了运镜控制描述的图生视频（科幻）场景中表现较差。

3.1 文生视频（武侠）

Prompt：一个快节奏的武侠短片片段。月光下的竹林，低角度仰拍两位古代侠客（一男一女）正在竹梢上飞跃交锋，他们剪影般的身影掠过月亮。随后是极速推进的特写，男侠客用两指夹住袭来的剑尖，剑气震落周围竹叶。画面充满动态模糊，树叶飞舞。

MiniMax Hailuo 02：整体指令遵循较好，视频符合物理规律；两指夹住袭来的剑尖的细节改为了用掌托住，细节方面略有不足。

Google Veo 3：整体指令遵循良好，视频符合物理规律；夹住剑的指令改为了男剑客单独持剑特写；场景切换相对突兀。

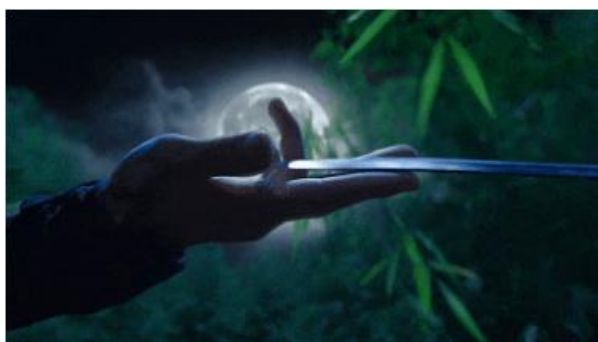
Runway Gen-4：整体指令遵循较好，视频符合物理规律；两指夹住袭来的剑尖的细节改为了夹住自己的剑；同样为男剑客单独特写；闪光特效相对出戏。

字节 Seedance 1.0：整体指令遵循尚可；人物略显僵硬，且无动作浮空，对物理规律的遵循不足；打斗特写无夹剑动作，且剑身存在消失、分裂等问题。

Sora 2：整体指令遵循较好，视频符合物理规律；两指夹住袭来的剑尖的细节改为了用手撑住，细节方面略有不足。

本轮测试中 Sora 2 视频生成水平基本与其他模型中表现最好的 MiniMax Hailuo 02 一致。

图 16：MiniMax Hailuo02 文生视频（武侠）效果



资料来源：久谦，中信建投

图 17：Google Veo 3 文生视频（武侠）效果



资料来源：久谦，中信建投

图 18: Runway Gen-4 文生视频（武侠）效果



资料来源：久谦，中信建投

图 19: 字节 Seedance 1.0 文生视频（武侠）效果



资料来源：久谦，中信建投

图 20: Sora 2 文生视频（武侠）效果



数据来源：Sora APP，中信建投

3.2 文生视频（悬疑）

Prompt: 模仿经典悬疑片的紧凑剪辑和特写镜头。场景设定在一个雨夜的废旧图书馆内，一名浑身湿透的情报员（男，30 岁左右，神情警惕）背靠书架剧烈喘息，突然，他猛地将一本厚实的古籍从书架中抽出。几乎同时，一名追踪者（女，身手矫健，戴战术手套）从阴影中迅速冲出，一把将古籍按回原处，用手臂将他死死压制在书架上，书页因撞击而散落。

快手可灵 2.1: 整体指令遵循良好，视频符合物理规律；雨夜环境体现的不明显；人物关系错误，二者从敌对关系被改为合作关系；未能生成“抽出书籍”动作，但“将书籍按回原处”、“书页散落”等剧情执行较好。

字节 Seedance 1.0: 整体指令遵循一般，其中前半程相对还原，后半段人物动作衔接混乱，明显不符合物理规律。

阿里 Wan2.2: 整体指令遵循尚可，视频符合物理规律；雨夜环境搭建合理，主体完成度高；未完成“抽出书籍”、“按回原处”等剧情。

Runway Gen-4: 整体指令遵循较好，视频符合物理规律，但动作表现力明显不足；“抽出书籍”、“书页散落”场景执行较好，仅“将书籍按回原处”被改为书籍掉落。

Google Veo3: 整体指令遵循一般，视频符合物理规律；场景构建较好，但仅生成了一个男性角色完成“抽出书籍”、“将书籍按回原处”、“书页散落”等剧情。

MiniMax Hailuo02: 整体指令遵循一般，视频符合物理规律，与 Google Veo3 类似，场景构建较好，但仅生成了一个男性角色执行动作。

生数科技 Vidu Q1: 整体指令遵循一般，人物瞬移且形态混乱，物理规则明显出错。

Sora 2: 整体指令遵循较好，视频符合物理规律，但表现力略有不足；较好执行了“抽出书籍”、“将书籍按回原处”等动作，但未能生成“书页散落”场景。

本轮测试中 Sora 2 和快手可灵 2.1、阿里 Wan2.2 共同位于第一梯队水平。其中快手可灵 2.1 在指令遵循方面稍弱，但视频质感和更强；Sora 2 和阿里 Wan2.2 则在指令遵循方面较好，动作表现力稍弱。

图 21：快手可灵 2.1 文生视频（悬疑）效果



资料来源：久谦，中信建投

图 22：字节 Seedance 1.0 文生视频（悬疑）效果



资料来源：久谦，中信建投

图 23：阿里 Wan2.2 文生视频（悬疑）效果



资料来源：久谦，中信建投

图 24：Runway Gen-4 文生视频（悬疑）效果



资料来源：久谦，中信建投

图 25: Google Veo 3 文生视频（悬疑）效果



资料来源：久谦，中信建投

图 26: MiniMax Hailuo02 文生视频（悬疑）效果



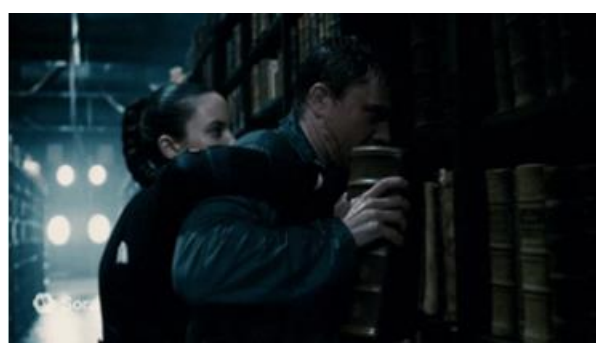
资料来源：久谦，中信建投

图 27: 生数科技 Vidu Q1 文生视频（悬疑）效果



资料来源：久谦，中信建投

图 28: Sora 2 文生视频（悬疑）效果



资料来源：Sora APP，中信建投

3.3 图生视频（喜剧、卡通）

Prompt: 以提供的卡通图片为起始帧，男人正低头看着手机，小心翼翼地走路，突然男人的脚精准地踩中香蕉皮，特写镜头捕捉到他脚底打滑的瞬间。下一刻，切换到慢动作：他的身体失去平衡，以一种违反物理定律的、极其夸张滑稽的方式在空中旋转了两圈，表情从专注瞬间变为极度惊恐，手中的手机也飞了出去，他四脚朝天地摔倒在地，眼冒金星。

Google Veo3: 整体指令遵循尚可；踩香蕉皮的动作较为刻意，且踩中后莫名跳跃；成功生成人物在空中旋转动作，但仅转了一圈，也生成了“手机飞出”的场景。

快手可灵 2.1: 整体指令遵循良好；对 Prompt 中“违反物理定律的旋转”理解产生了偏差，踩中香蕉皮后人物变形扭曲，不再符合物理规律。

爱诗科技 PixVerse V5: 整体指令遵循一般；踩香蕉皮和摔倒的动作均较为刻意，且未生成“手机飞出”的场景；角色对电话大骂的生成内容也与用户意图偏离。

生数科技 Vidu Q1: 整体指令遵循一般，视频符合物理规律；未执行踩香蕉皮动作，而是突然摔倒。

Sora 2: 整体指令遵循较好；踩香蕉皮的动作较为刻意；成功生成人物在空中旋转动作，但仅转了一圈，且旋转过程中人物腿部有所变形；成功生成了“手机飞出”的场景。

本轮测试中 Sora 2 表现超过其他模型水平。

图 29：Google Veo3 图生视频（喜剧、卡通）效果



资料来源：久谦，中信建投

图 30：快手可灵 2.1 图生视频（喜剧、卡通）效果



资料来源：久谦，中信建投

图 31：爱诗科技 PixVerse V5 图生视频（喜剧、卡通）效果



资料来源：久谦，中信建投

图 32：生数科技 Vidu Q1 图生视频（喜剧、卡通）效果



资料来源：久谦，中信建投

图 33：Sora 2 图生视频（喜剧、卡通）效果



数据来源：Sora APP，中信建投

3.4 图生视频（科幻）

Prompt：让海报中的角色“活过来”。起始画面是静态海报，随后模拟一个缓慢的无人机视角后退拉升镜头，揭示出更广阔的末日废墟场景。同时，海报中的主角开始缓缓转过头，看向镜头方向，眼神中充满希望。风中，他的衣角和远处的烟雾轻轻飘动。色调保持苍凉但壮丽，具有史诗感。

MiniMax Hailuo02：整体指令遵循较好；运镜平稳，但后退拉升镜头的幅度略有不足；“衣角飘动”的场景生成较好，但未能生成“远处的烟雾”，而是生产了地上的飞尘。

Google Veo3：整体指令遵循较好，所有元素均囊括其中，尤其“衣角和远处烟雾飘动”的场景完美生成，质感强；美中不足的是把后退拉升镜头弄成了向前推镜头。

爱诗科技 PixVerse V5：整体指令遵循较好，所有元素均囊括其中，仅把“远处的烟雾”改为了“主角身边的烟雾”；同样把后退拉升镜头弄成了向前推镜头。

快手可灵 2.1：整体指令遵循较好，所有元素均囊括其中，尤其末日废墟场景表现不错，但整体场景类似贴图缩放，缺乏质感。

生数科技 Vidu Q1：整体指令遵循良好；未生成“烟雾”元素；“无人机视角后退拉升镜头”被理解为“镜头跟随人物向后移动”，违反物理规律。

Sora 2：整体指令遵循一般，不仅未能生成“衣角和远处烟雾飘动”，镜头也只是水平移动。

本轮测试中 Sora 2 相对较弱，仅优于违反物理规律的生数科技 Vidu Q1。

图 34：MiniMax Hailuo02 图生视频（科幻）效果



资料来源：久谦，中信建投

图 35：Google Veo3 图生视频（科幻）效果



资料来源：久谦，中信建投

图 36：爱诗科技 PixVerse V5 图生视频（科幻）效果



资料来源：久谦，中信建投

图 37：快手可灵 2.1 图生视频（科幻）效果



资料来源：久谦，中信建投

图 38：生数科技 Vidu Q1 图生视频（科幻）效果



资料来源：久谦，中信建投

图 39：Sora 2 图生视频（科幻）效果



资料来源：Sora App，中信建投

四、AI 视频生成市场空间有几何？

AI 视频生成技术作为颠覆性的生产力工具，商业化潜力巨大。我们将分别从 P 端（专业创作者）、B 端（企业级应用）和 C 端（大众消费）三个维度，对 AI 视频生成的中期（3 年）和长期（5 年）市场空间进行测算。

1）全球 P 端创作者经济市场中期 262 亿元空间，长期 888 亿元空间（按美元：人民币汇率 7：1 计算）。

P 端用户付费意愿强，是 AI 视频工具最直接的变现市场，我们沿用“创作者人数 × AI 产品渗透率 × ARPU”的公式对国内和全球市场进行测算。

海外：P 端市场空间中期 36 亿美元空间，长期 120 亿美元空间。

I）创作者人数：据 Research Nester 数据，截至 2023 年 5 月海外内容创作者数量超过 2.08 亿，其中巴西数量最多达 1.05 亿，美国其次为 8650 万。通过 TikTok 数据进行验证（Thunderbit），2025 年 TikTok MAU 15.9 亿，83% 的用户发过视频，2.08 亿对应 15.8% 的核心内容创作者，相对合理。综上我们假定中期海外创作者人数为 2 亿，随着数字经济发展及 AI 工具普及创作者人数长期提升至 5 亿。**II）AI 产品渗透率：**2025 年 4 月 ChatGPT 5 亿周活（IT 之家），付费用户突破 2000 万（Information），对应付费渗透率 4%。综上我们假定 AI 视频工具产品中期付费渗透率为 5%，随着数字经济发展及 AI 工具长期保守提升至 20%（当前全球 90% 的创意专业人士都在使用 Adobe Photoshop）。**III）ARPU：**快手可灵海外版 Standard/Pro/Premier/Ultra 会员（不考虑优惠）每月分别 10/37/92/180 美元，按照 6：2：1：1 加权计算为 40.6 美元/月。考虑折扣下我们假定 AI 视频工具产品中期订阅价格为 30 美元/月，长期由于模型优化和推理成本降低下降至 10 美元/月。**IV）综上所述海外 P 端市场中期 36 亿美元空间，长期 120 亿美元空间。**

国内：P 端市场空间中期 10 亿元空间，长期 48 亿元空间。

I）创作者人数：据《2025 年网络视听内容创作者白皮书》数据，截至 2024 年底，短视频创作者账号规模也已达 16.2 亿，全国每日短视频上传量突破 1.3 亿条。国内头部短视频平台包括抖音、快手、视频号、小红书、哔哩哔哩等，参考海外我们假定其中 15% 为核心内容创作者，且人均账号为 3 个，对应 0.81 亿创作者。综上我们假定中期国内创作者人数为 0.8 亿，随着数字经济发展及 AI 工具普及创作者人数长期提升至 2 亿。**II）AI 产品渗透率：**考虑国内付费意愿低于海外，我们假定 AI 视频工具产品中期付费渗透率为 1%，随着数字经济发展及 AI 工具长期保守提升至 5%。**III）ARPU：**快手可灵黄金/铂金/钻石/黑金会员（不考虑优惠）每月分别 58/234/586/1314 元，按照 8：1：0.5：0.5 加权计算为 138.4 元/月。考虑折扣下我们假定 AI 视频工具产品中期订阅价格为 100 元/月，长期由于模型优化和推理成本降低下降至 40 元/月。**IV）综上所述国内 P 端市场中期 9.6 亿元空间，长期 48 亿元空间。**

表 2:AI 视频生成产品 P 端市场规模测算

	中期（2028 年）		长期（2030 年）	
	海外	国内	海外	国内
创作者人数	2 亿	0.8 亿	5 亿	2 亿
AI 产品渗透率	5%	1%	20%	5%
ARPU	30 美元/月	100 元/月	10 美元/月	40 元/月
市场规模	36 亿美元	9.6 亿元	120 亿美元	48 亿元
合计	262 亿元		888 亿元	

资料来源：Research Nester，Thunderbit，IT 之家，The Information，可灵，《2025 年网络视听内容创作者白皮书》，中国日报网，中信建投；注：美元：人民币汇量按 7：1 计算

2) 全球 B 端垂直场景市场中期 501 亿元空间，长期 666 亿元空间（按美元：人民币汇率 7：1 计算）。

B 端市场 AI 视频生成应用最广的领域，核心在于对传统制作流程的成本替代，可预见的应用场景包括专业影视（电影/游戏 CG）、广告、短剧、培训（教育/企业）等。我们按照“传统制造成本 $\times$ AI 渗透率 $\times$ AI 成本优化”进行测算。

关键假设：由于目前相关数据不足，假设较为主观。

I) AI 渗透率：中期头部公司和工作室开始试点和在非核心项目中使用，长期 AI 成为行业标准流程，大部分制作公司采纳 AI 工具链。按照专业性要求不同，我们假定中期影视/广告/短剧/培训等行业 AI 渗透率分别达到 5%/10%/30%/30%，长期渗透率分别达到 15%/30%/60%/60%。

II) AI 成本优化：中国日报网报道，Chinagoods AI 创新项目组工程师代表楼晗啸在社媒营销获客 AI 峰会上指出，以往一条 TVC 需数十万成本、数月周期，如今通过 AI 生成，成本不及传统方式十分之一；新华网报道，国际知名视效指导姚骏仪花费 330 元即制作出科幻短片《归途》，而传统制作模式可能需要百万级别预算。尽管上述案例降本幅度较大，但考虑到中期仅预演、素材生成、特效合成等少部分场景 AI 应用空间大，长期 AI 才会逐步渗透至替代实拍、数字演员等场景。结合不同场景的专业性要求，我们假定中期影视/广告/电商视频/短剧/培训等行业使用 AI 工具的成本为传统方式的 70%/50%/30%/20%，长期成本降低为 30%/20%/10%/5%。

各细分领域传统制造成本：

专业影视（电影/游戏 CG）：当前市场规模约 2430 亿人民币，按 5% CAGR 估计，中期和长期市场规模分别 2810 亿元/3100 亿元。**I) 全球电影市场：**2024 年全球电影票房 280 亿美元（猫眼）；而票房总收入达到制作成本的 3 倍才有可能收回成本（澎湃新闻）；考虑到电影发行行业毛利率约 30%，则电影制作成本约占票房的 23%，对应 65 亿美元，约 460 亿人民币。**II) 全球游戏市场：**据 Newzoo 报告，2024 年全球游戏市场规模约 1877 亿美元。其中，3A 大作的开发成本中，美术与 CG 的成本占比极高，通常达到 25%-30%（据 kotaku，《漫威蜘蛛侠 2》的预算为 3.15 亿美元，其中仅 314 分钟的过场动画成本就超过了 4000 万美元）。考虑小型独立游戏对美术需求较低，我们假定全行业 CG 相关成本保守估计为 15%，对应 282 亿美元，约 1970 亿人民币。

视频广告：当前市场规模约 4690 亿人民币，按 10% CAGR 估计，中期和长期市场规模分别 6240 亿元/7550 亿元。据 Business Research Insight，全球广告视频产量市场规模在 2024 年为 670.4 亿美元，对应约 4690 亿人民币。

短剧：当前市场规模为 300 亿人民币，按 25% CAGR 估计，中期和长期市场规模分别 590 亿元/920 亿元。据 DataEye 研究院，2024 年中国微短剧市场规模达 504 亿，预计 2025 年将超过 680 亿，2027 年将突破 1000 亿。据南方都市报访谈创作人，短剧制造成本占总成本 60%-70%，按 60% 估计短剧制造市场约 300 亿元。

培训（教育/企业视频制作）：当前市场规模为 410 亿人民币，按 5% CAGR 估计，中期和长期市场规模分别 470 亿元/520 亿元。**I) 在线教育市场：**据 Business Research Insight，2024 年全球在线教育市场规模约 1782 亿美元，由于缺乏数据支撑，我们保守估计视频课程内容开发费用占比约为 3%，对应市场空间 53 亿美元，合 370 亿人民币。**2) 企业培训市场：**据 Business Research Insight，2024 年全球企业培训市场规模约 192 亿美元，我们同样保守估计数字化学习（e-learning）视频制作成本占比约为 3%，对应市场空间 6 亿美元，合 40 亿人民币。

综上所述，专业影视、短剧、广告、培训场景 AI 视频市场中期 501 亿元空间，长期 666 亿元空间。

表 3: AI 视频生成产品 B 端市场规模测算

	中期（2028 年）				长期（2030 年）			
细分领域	专业影视	视频广告	短剧	培训	专业影视	视频广告	短剧	培训
传统制造成本（亿元）	2810	6240	590	470	3100	7550	920	520
AI 渗透率	5%	10%	30%	40%	15%	30%	60%	70%
使用 AI 工具成本/原始成本	70%	50%	30%	20%	30%	20%	10%	5%
市场空间（亿元）	98	312	53	38	140	453	55	18
合计	501				666			

资料来源：中国日报网，新华网，猫眼，澎湃新闻，Newzoo，游民星空，kotaku，36kr、DataEye 研究院，Business Research Insight，中信建投；注：美元：人民币汇率按 7：1 计算

3) C 端市场：流量经济的全新入口

基于 Sora APP 的当前情况，我们认为其 C 端模式极大概率不会直接收费，而是效仿 TikTok 等短视频平台通过 AI 原生内容社区的定位构建庞大的用户流量池，最终通过广告变现。这类商业模式并不属于 AI 视频工具的市场空间，而是提供 B 端市场新的增量（上述 B 端市场测算未考虑 AI 原生内容社区带来的广告增量）。

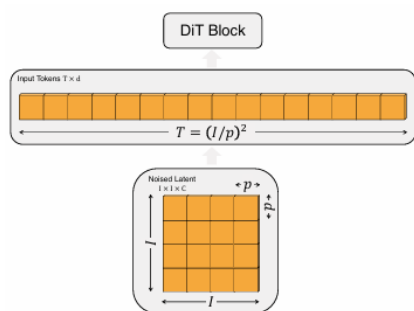
综合上述分析，AI 视频生成市场中中期市场空间为 763 亿元，长期市场空间为 1554 亿元。

五、Sora 2 的 Token 成本有多少？

1 分钟高清视频约 1M tokens。参考 DiT 论文，256x256 的图片会被划分为 32x32 个 patch，对应一个 patch 为 8x8 的像素块。我们假设 1920x1080 分辨率的高清图像经过下采样得到 512x256 大小的图片，对应包括 2048 个 patch。1s 高清视频往往 30-60 帧，同时考虑 Sora 在训练和推理过程中进行了压缩，我们假定压缩后的 1s 视频为 8 帧，则一分钟的视频共 480 帧，对应 983040 个 patch。

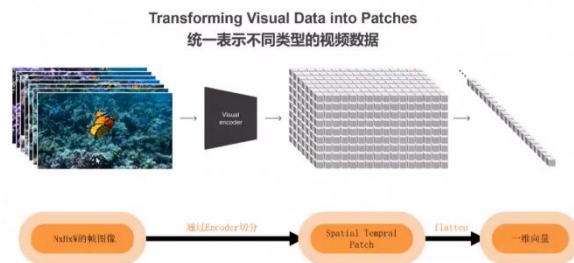
DiT 把视频切成 $m \times n \times t$ 三维的 patch（仍是 RGB 像素空间）阵列后，经过空间映射会得到 latent 空间中的一维序列，一个 patch 对应的一个 latent 元素，等价于 LLM 中的一个 token（经过字符空间映射过来的 token 空间基本单元），则估计 1 分钟的视频约 1M tokens。

图 40：DiT Patch 的拆分



资料来源：《Scalable Diffusion Models with Transformers》，中信建投

图 41：Patch 到向量的映射过程



资料来源：CSDN，中信建投

预计 Sora 2 模型推理 1 分钟 1080 高清视频的总计算量为 1.8E+7 TFLOPs。根据 OpenAI 的《Scaling Laws for Neural Language Models》，训练 Transformer 模型的理论计算量为总算力需求 $C \approx 6ND$ ，推理的总算力

请务必阅读正文之后的免责条款和声明。

需求为 $C \approx 2ND$ 。其中, N 为模型参数量大小, D 为训练数据量大小。需要注意的是, 该公式成立的前提是 $d_{model} \gg n_{ctx}/12$ 。其中, d_{model} 为残差流维度, n_{ctx} 为输入文本的 token 数。由于视频训练的输入 tokens 不再符合上述前提, 则实际公式引述昆仑万维 AI Infra 负责人成诚的测算应为 $C \approx 90ND$, 推理的算力需求公式为 $C \approx 30ND$ 。

Diffusion 架构基于随机的 noise latent 矩阵按照多个时间步迭代生成, 我们假设 Sora 在实际推理时, 需要 20 个 step 生成视频。

按照上述公式计算, 假定 Sora 2/Sora 2 Pro 模型为 20B/30B, 则其推理 1 分钟高清视频 (1M tokens) 的计算量为 1.18E+7 TFLOPs/1.77E+7 TFLOPs。

图 42: Transformer 架构 LLM 每 token 需要的计算量是 6 倍的模型大小

Operation	Parameters	FLOPs per Token
Embed	$(n_{vocab} + n_{ctx}) d_{model}$	$4d_{model}$
Attention: QKV	$n_{layer} d_{model} 3d_{attn}$	$2n_{layer} d_{model} 3d_{attn}$
Attention: Mask	—	$2n_{layer} n_{ctx} d_{attn}$
Attention: Project	$n_{layer} d_{attn} d_{model}$	$2n_{layer} d_{attn} d_{embd}$
Feedforward	$n_{layer} 2d_{model} d_{ff}$	$2n_{layer} 2d_{model} d_{ff}$
De-embed	—	$2d_{model} n_{vocab}$
Total (Non-Embedding)	$N = 2d_{model} n_{layer} (2d_{attn} + d_{ff})$	$C_{forward} = 2N + 2n_{layer} n_{ctx} d_{attn}$

Table 1 Parameter counts and compute (forward pass) estimates for a Transformer model. Sub-leading terms such as nonlinearities, biases, and layer normalization are omitted.

For contexts and models with $d_{model} > n_{ctx}/12$, the context-dependent computational cost per token is a relatively small fraction of the total compute. Since we primarily study models where $d_{model} \gg n_{ctx}/12$, we do not include context-dependent terms in our training compute estimate. Accounting for the backwards pass (approximately twice the compute as the forwards pass), we then define the estimated non-embedding compute as $C \approx 6N$ floating point operators per training token.

数据来源:《Scaling Laws for Neural Language Models》, 中信建投

保守估计 Sora APP 将带来每天 1400 万美元的推理成本, 年化 51.2 亿美元。2025 年 10 月 7 日, OpenAI 于 2025 年开发者大会上公布了 Sora 2 的 API 价格, Sora 2/Sora 2 Pro 生成 720p 分辨率的视频每秒分别 \$0.1/\$0.3; Sora 2 Pro 生成 1024p 分辨率的视频每秒 \$0.5。考虑到 1080p 的分辨率为 360p 的 9 倍, 为 720p 的 2.25 倍, 为 1024p 的 1.13 倍, 则 1s 360p/720p/1024p 的 tokens 消耗量分别为 1.82K/7.28K/14.50K。同时据可灵 AI 称已经在推理算力层面实现了毛利率的打正, 我们假定视频生成模型的 API 价格即为推理成本, 则 Sora 2/Sora 2 Pro 模型生成 1 秒 720p 分辨率的成本为每百万 tokens 13.73/41.20 美元, Sora 2 Pro 模型生成 1 秒 1024p 分辨率的成本为每百万 tokens 34.49 美元。

按上述假设, Sora 2 模型推理 10 秒钟 360p 视频的计算量为 2.18E+5 TFLOPs; Sora 2/Sora 2 Pro 模型推理 10 秒钟 720p 视频的计算量为 8.74E+5 TFLOPs/1.31E+6 TFLOPs; Sora 2 Pro 模型推理 10 秒钟 1024p 视频的计算量为 2.61E+6 TFLOPs。

按 H100 单卡算力 1979TFLOPs (FP16) 计算, 假设 Sora 2/Sora 2 Pro 模型的 GPU 利用率为 50% (DiT 是

Encoder-Only Transformer 架构，推理时会一次性输出全部长度的 token，为 Compute Bound 场景，对 GPU 的利用率更高）。则 Sora 2 模型推理 10 秒钟 360p 视频的需要 221 H100-秒；Sora 2/Sora 2 Pro 模型推理 10 秒钟 720p 视频的需要 883/1325 H100-秒；Sora 2 Pro 模型推理 10 秒钟 1024p 视频的需要 2637 H100-秒。

据 OpenAI 在 2025 年开发者大会上公布的最新数据，ChatGPT 周活跃用户已达到 8 亿，随着邀请码裂变，我们假定 ChatGPT 用户最终都会使用 Sora APP，且每周使用一次 Sora APP 生成 10 秒钟 360p 视频（实际视频生成功能使用频率将会显著高于 Chatbot 类产品，且未考虑 Plus 和 Pro 用户生成更高质量，更长时间视频带来的额外推理需求），则对应每天 $7.01E+7$ H100-小时推理算力需求，按 H100 每小时 2 美元租赁价格，合计 1400 万美元推理成本，年化 51.2 亿美元。

表 4: Sora 2 视频生成算力需求测算

模型	Sora 2				Sora 2 Pro			
总参数量 (N, B)	20				30			
激活参数量 (N, B)	20				30			
分辨率	360p	720p	1024p	1080p	360p	720p	1024p	1080p
1 分钟视频输出 token 长度 (D, K)	109.23	436.91	869.93	983.04	109.23	436.91	869.93	983.04
10 秒钟视频输出 token 长度 (D, K)	18.20	72.82	144.99	163.84	18.20	72.82	144.99	163.84
总浮点计算公式	C=30ND				C=30ND			
总推理迭代次数	20				20			
总推理算力需求 (TFLOPs)	2.18E+05	8.74E+05	1.74E+06	1.97E+06	3.28E+05	1.31E+06	2.61E+06	2.95E+06
H100 单卡算力 (TFLOPs)					1979			
GPU 利用率					50%			
算力需求 (H100-秒)	221	883	1758	1987	331	1325	2637	2980

资料来源：《Scaling Laws for Neural Language Models》，Nvidia，中信建投

六、投资建议

Sora 将从 AI 应用和算力两个维度促进 AI 产业链发展。

AI 应用：Sora 2 三天登顶美国 iOS 应用榜，直接反映了“为产品定义功能，而非为模型定义产品”思路的重要性。考虑到 ChatGPT 周活跃用户持续高增（7 月底 7 亿周活，9 月底 8 亿周活），AI 应用赛道预计将保持景气度。

1) 重视以阿里为代表的国内互联网大厂 Capex 投资和生态圈玩家，包括国产芯片链企业、isv 厂商等。

2) Pre-AI 的收入率先落地。国内企业数据治理相对海外落后，AI 应用落地需要更多前期准备工作。OA+ERP 作为集成企业数据的平台入口直接受益，推荐泛微网络、金蝶国际、用友网络、鼎捷数智；Pre-AI 环节推荐海天瑞声、达梦数据、深信服、汉得信息等。

3) 部分细分垂直场景 AI 收入更快。其中创收关注医疗、教育、市场营销、Science 侧产品，推荐拓尔思、税友股份、视源股份、京北方、国能日新、晶泰控股等；降本关注办公、AI-coding、多模态生成，推荐万兴科技；办公场景，推荐金山办公；场景数字化，推荐海康威视、大华股份。

算力：9 月 22 日，OpenAI 和 Nvidia 达成合作，OpenAI 将部署包括 NVIDIA Vera Rubin 平台在内的至少 10 吉瓦的 NVIDIA 系统，用于 OpenAI 的下一代 AI 基础设施。NVIDIA 计划向 OpenAI 投资高达 1000 亿美元，随着每吉瓦的部署逐步到位。10 月 6 日，OpenAI 与 AMD 达成合作，OpenAI 将在未来数年部署高达 6GW 的 AMD Instinct GPU。根据协议，首批 1GW 设备将于 2026 年投入使用。AMD 已向 OpenAI 发行最高 1.6 亿股认股

权证，行权条件与芯片部署进度及股价里程碑挂钩。如果 OpenAI 全额行权，基于 AMD 当前流通股总数计算，它可能获得 AMD 约 10% 的股权。而根据我们测算，保守估计 Sora APP 将带来每天 1400 万美元的推理成本，年化 51.2 亿美元，庞大的推理算力需求与上述合作相契合。

1) 本地推理逐步起量，利好服务和云厂商。推荐浪潮信息、中科曙光、海光信息、深信服、神州数码、联想集团、工业富联等。

2) 推荐关注端侧 AI 和相关产业链，推荐虹软科技、中科创达等。

风险分析

(1) AI 产业商业化落地不及预期：目前各环节 AI 产品的商业化模式尚处于探索阶段，如果各环节产品的推进节奏不及预期，或对相关企业业绩造成不利影响；(2) 市场竞争风险：海外 AI 厂商凭借先发优势，以及较强的技术积累，在竞争中处于优势地位，如果国内 AI 厂商技术迭代不及预期，经营状况或将受到影响；同时，目前国内已有众多企业投入 AI 产品研发，后续可能存在同质化竞争风险，进而影响相关企业的收入；(3) 政策风险：AI 技术的发展直接受各国政策和监管影响。随着 AI 在各个领域的渗透，政府可能会进一步出台相应的监管政策以规范其发展。如果企业未能及时适应和遵守相关政策，可能面临相应处罚，甚至被迫调整业务策略。此外，政策的不确定性也可能导致企业战略规划和投资决策的错误，增加运营的不确定性；(4) 地缘政治风险：在全球地缘政治环境的波动下，尤其美国对中国的出口限制或将直接影响国内企业算力芯片的获取，进而影响其产品研发和市场竞争能力。同时，地缘政治风险也可能导致 AI 产品开拓海外市场面临障碍，影响相关企业的营收情况。

分析师介绍

应瑛

中信建投证券计算机行业首席分析师，伦敦国王学院硕士，5年计算机行业研究经验。2021年加入中信建投，深入覆盖医疗信息化、工业软件、云计算、网络安全等细分领域。

研究助理

李楚涵

lichuhan@csc.com.cn

评级说明

投资评级标准		评级	说明
报告中投资建议涉及的评级标准为报告发布日后6个月内的相对市场表现,也即报告发布日后的6个月内公司股价(或行业指数)相对同期相关证券市场代表性指数的涨跌幅作为基准。A股市场以沪深300指数作为基准;新三板市场以三板成指为基准;香港市场以恒生指数作为基准;美国市场以标普500指数为基准。	股票评级	买入	相对涨幅 15%以上
		增持	相对涨幅 5%—15%
		中性	相对涨幅-5%—5%之间
		减持	相对跌幅 5%—15%
		卖出	相对跌幅 15%以上
	行业评级	强于大市	相对涨幅 10%以上
		中性	相对涨幅-10-10%之间
		弱于大市	相对跌幅 10%以上

分析师声明

本报告署名分析师在此声明:(i)以勤勉的职业态度、专业审慎的研究方法,使用合法合规的信息,独立、客观地出具本报告,结论不受任何第三方的授意或影响。(ii)本人不曾因,不因,也将不会因本报告中的具体推荐意见或观点而直接或间接收到任何形式的补偿。

法律主体说明

本报告由中信建投证券股份有限公司及/或其附属机构(以下合称“中信建投”)制作,由中信建投证券股份有限公司在中华人民共和国(仅为本报告目的,不包括香港、澳门、台湾)提供。中信建投证券股份有限公司具有中国证监会许可的投资咨询业务资格,本报告署名分析师所持中国证券业协会授予的证券投资咨询执业资格证书编号已披露在报告首页。

在遵守适用的法律法规情况下,本报告亦可能由中信建投(国际)证券有限公司在香港提供。本报告作者所持香港证监会牌照的中央编号已披露在报告首页。

一般性声明

本报告由中信建投制作。发送本报告不构成任何合同或承诺的基础,不因接收者收到本报告而视其为中信建投客户。

本报告的信息均来源于中信建投认为可靠的公开资料,但中信建投对这些信息的准确性及完整性不作任何保证。本报告所载观点、评估和预测仅反映本报告出具日该分析师的判断,该等观点、评估和预测可能在不发出通知的情况下有所变更,亦有可能因使用不同假设和标准或者采用不同分析方法而与中信建投其他部门、人员口头或书面表达的意见不同或相反。本报告所引证券或其他金融工具的过往业绩不代表其未来表现。报告中所含任何具有预测性质的内容皆基于相应的假设条件,而任何假设条件都可能随时发生变化并影响实际投资收益。中信建投不承诺、不保证本报告所含具有预测性质的内容必然得以实现。

本报告内容的全部或部分均不构成投资建议。本报告所包含的观点、建议并未考虑报告接收人在财务状况、投资目的、风险偏好等方面的具体情况,报告接收者应当独立评估本报告所含信息,基于自身投资目标、需求、市场机会、风险及其他因素自主做出决策并自行承担投资风险。中信建投建议所有投资者应就任何潜在投资向其税务、会计或法律顾问咨询。不论报告接收者是否根据本报告做出投资决策,中信建投都不对该等投资决策提供任何形式的担保,亦不以任何形式分享投资收益或者分担投资损失。中信建投不对使用本报告所产生的任何直接或间接损失承担责任。

在法律法规及监管规定允许的范围内,中信建投可能持有并交易本报告中所提公司的股份或其他财产权益,也可能在过去12个月、目前或者将来为本报中所提公司提供或者争取为其提供投资银行、做市交易、财务顾问或其他金融服务。本报告内容真实、准确、完整地反映了署名分析师的观点,分析师的薪酬无论过去、现在或未来都不会直接或间接与其所撰写报告中的具体观点相联系,分析师亦不会因撰写本报告而获取不当利益。

本报告为中信建投所有。未经中信建投事先书面许可,任何机构和/或个人不得以任何形式转发、翻版、复制、发布或引用本报告全部或部分内容,亦不得从未经中信建投书面授权的任何机构、个人或其运营的媒体平台接收、翻版、复制或引用本报告全部或部分内容。版权所有,违者必究。

中信建投证券研究发展部

北京
朝阳区景辉街16号院1号楼18层
电话:(8610) 56135088
联系人:李祉瑶
邮箱:lizhiyao@csc.com.cn

上海
上海浦东新区浦东南路528号南塔2103室
电话:(8621) 6882-1600
联系人:翁起帆
邮箱:wengqifan@csc.com.cn

深圳
福田区福中三路与鹏程一路交汇广电金融中心35楼
电话:(86755) 8252-1369
联系人:曹莹
邮箱:caoying@csc.com.cn

中信建投(国际)

香港
中环交易广场2期18楼
电话:(852) 3465-5600
联系人:刘泓麟
邮箱:charleneliu@csci.hk